UNIVERSITÉ DE MONTRÉAL

CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, STATION A, MONTRÉAL, QUÉBEC, H3C 3J7

ESTIMATING ORIGIN-DESTINATION MATRICES FROM TRAFFIC COUNTS: A SIMPLE LINEAR INTERCITY MODEL FOR QUEBEC

by

Marc Gaudry* and Louis Lamarre**

* Département des sciences économiques et Centre de recherche sur les transports, Université de Montréal.

** Lavalin Inc., Montréal.

This research was partly subsidized by the Research and Development Centre of Transport Canada and by the F.C.A.C. program of the *Ministère de l'éducation du Québec*.

Université de Montréal

Publication # 105 - Centre de recherche sur les transports
Cahier # 7809 - Département de sciences économiques

Août 1978

Dépôt légal - 4e trimestre 1978

ABSTRACT

This paper presents a linear model to estimate an intercity origin-destination matrix from volume counts on network links. It is argued that the use of a single impedance term in the assignment procedure and the taking into account of heteroskedasticity of the residuals circumvent some of the difficulties presented by O-D estimation problems. Indirect evidence from residuals of the linear model and from a gravity format suggest that significant non-linearities are present in the data set and could be isolated if the present state of techniques permitted to distinguish between the functional form of a model and heteroskedasticity of the residuals.

CONTENTS

		Page
1.	INTRODUCTION	1
2.	MODEL SPECIFICATION	2
	A. The joint determination of costs and of their weights	4
	B. Heteroskedasticity and the functional form of the demand function	4
	C. A linear specification with homoskedastic errors	6
3.	RESULTS	8
	A. General comments	8
	B. Summary comparison with gravity format results	8
4.	CONCLUSION	10
ACKNOWLEDGEMENTS		11
REFERENCES		12

Figure 1. Quebec network in 1972 with numbered nodes and links..... 3

Table 1. Results for linear specification..... 9

Table 2. Travel time demand elasticities..... 9

1. INTRODUCTION

The idea of obviating the need to conduct origin-destination surveys by estimating origin-destination (henceforth O-D) matrices from counts made on the links of a network of transit lines, streets or roads is relatively recent. Working over the period 1969-1971, a group of City of Montreal planners (Arbour *et al* 1969; 1971a; 1971b) have summarized in technical reports a model which estimated O-D matrices from link counts for *total* trips by all modes and for *transit* trips among 60 large zones of the Montreal metropolitan area. The matrices were estimated from distinct sequences of linear trip generation and gravity distribution equations; the procedure for the transit matrix used three different distribution equation variants simultaneously. The purpose of that unpublicized work was the simulation of ridership for proposed major new metro lines. Independently, Wills (1971), in an unpublished thesis, obtained intercity O-D matrices of British Columbia for various time periods using a gravity model. The first generally available analysis was apparently made by Low (1972). Overgaard (1972), Jentsen and Nielsen (1973), described in Bendtsen (1974) or in OECD (1974), and Robillard and Trahan (1973) also made early contributions.

All of these analyses and subsequent papers by Robillard (1974), Wills (1974), Branston *et al* (1975), the Danish Road Directorate, reported by Holm *et al* (1976), and Högberg (1976) are centered on the specification and use of a demand model. This is also true of Debaille's (1977) related procedure. By contrast, Nguyen (1977a; 1977b; 1978) has developed an approach based on equilibrium assignment and Van Zuylen (1978) and Willumsen (1978a; 1978b) have specified models based on entropy.

The studies based on demand models have all estimated the parameters of the gravity multiplicative model. This paper will present results based on a simple linear model which takes into account heteroskedasticity of the residuals; the "best" results will be summarily compared to those obtained with an unconstrained generation-distribution gravity model which does not take heteroskedasticity of the residuals into account. The data to be explained consist of two-way traffic counts on 37 links of a network of 23 nodes for Quebec in 1972. The network is shown in figure 1.

2. MODEL SPECIFICATION

Given p_{ij}^{s} , the proportion of trips from origin i to destination j which use network link s, the basic equations of the model are

$$F_s = \sum_i \sum_j p_{ij}^{s} \, T_{ij} \tag{1}$$

and

$$T_{ij} = f\left(X_{ij_1}, \ldots, X_{ij_K}\right) , \tag{2}$$

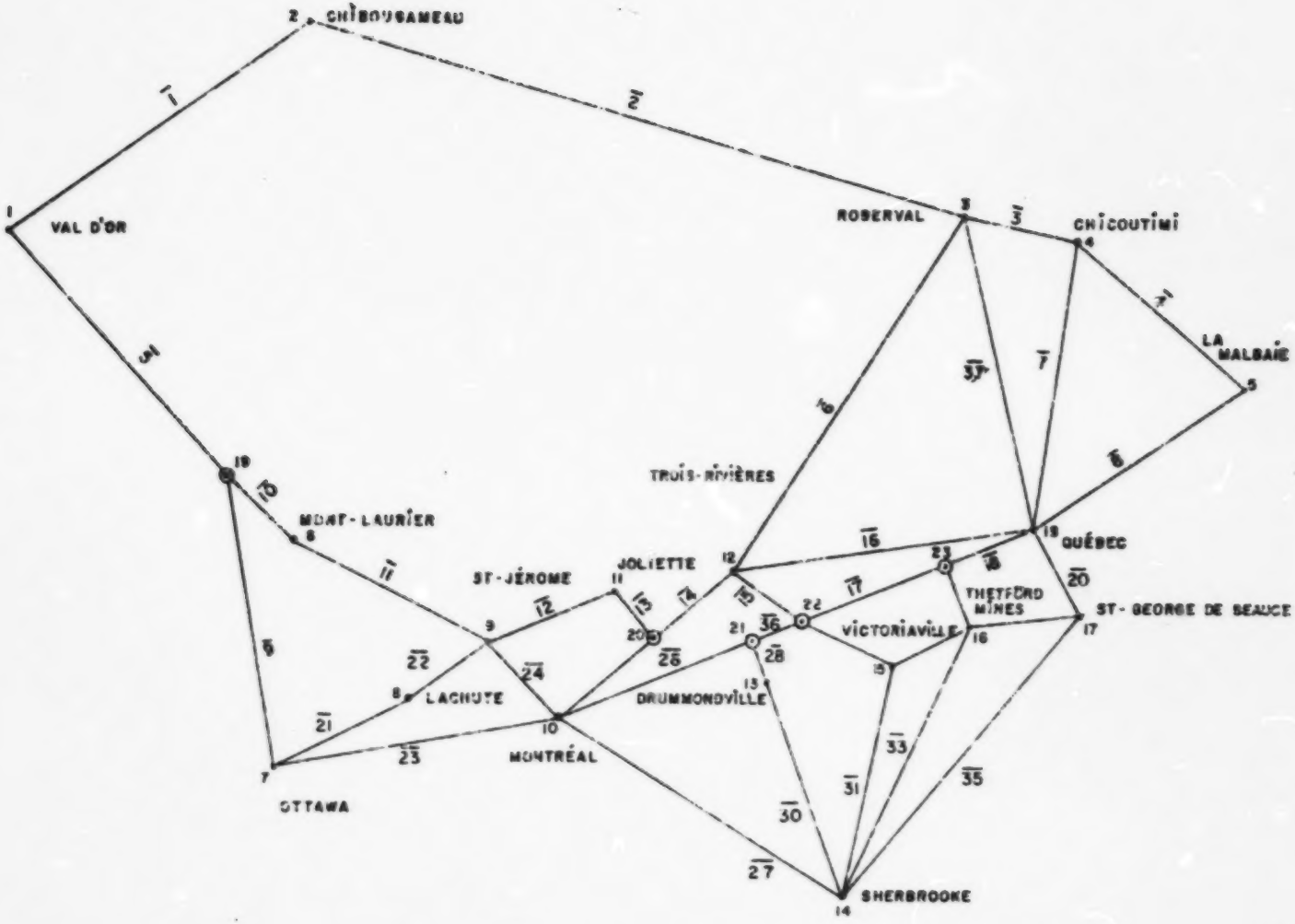
where

F_s = traffic volume on the s^{th} link of the network;

T_{ij} = number of trips from i to j;

X_{ij_k} = a function of the k^{th} activity level, socioeconomic indicator or (money *or* time) cost associated with origin i, destination j, or both of them.

Two problems should be mentioned before stating our exact specification.



QUEBEC NETWORK IN 1972 WITH NUMBERED NODES (x) AND LINKS ($\bar{x}$)

FIGURE 1

A. The joint determination of costs and of their weights

In an intercity network in which there are few competing itineraries, the p^s_{ij} in (1) can reasonably be obtained from an all-or-nothing shortest time path assignment procedure. Shortest generalized cost (time + money) assignment techniques raise the basic problem of the simultaneous determination of the time and cost (or distance) combinations to be used in the demand function for an estimation of their relative weights and of the relative weights to be used in the generalized cost assignment procedure. These problems of joint determination of time and distance matrices and of their relative valuation by travellers arise even in the absence of congestion and are discussed in more detail in Gaudry (1977). In this paper the p^s_{ij} are obtained from shortest time paths, except for a few modifications made by hand in the linear model in order to use extraneous information.

B. Heteroskedasticity and the functional form of the demand function

Specifying the functional form of (2) in a still more difficult problem. A general approach could use a Box-Cox (1964) transformation on all non-Boolean variables of (2) as follows

$$\frac{T^{\lambda}_{ij} - 1}{\lambda} = \beta_o + \sum_k \beta_k \left(\frac{X^{\lambda}_{ij_k} - 1}{\lambda} \right), \tag{3}$$

where, for any variable z considered,

$$\lim_{\lambda \to 0} \frac{z^{\lambda} - 1}{\lambda} = \ln z \ . \tag{4}$$

Form (3) includes the gravity model ($\lambda \to 0$) and the linear model ($\lambda = 1$) as special cases. Isolating T_{ij} in (3), and combining the result with (1) yields

$$T_s = \sum_i \sum_j p_{ij}^s \left\{ \lambda \left[\beta_o + \sum_k \beta_k \left(\frac{x_{ij_k}^{\lambda} - 1}{\lambda} \right) \right] + 1 \right\}^{1/\lambda} + e_s \quad . \qquad (5)$$

One could envisage calibrating the parameters λ, β_o, β_1,..., β_K included in (5) by minimizing the sum of squared errors on the links —— and this will be done later for $\lambda = 0$. Such a curve fitting procedure, however, provides no way of testing for the statistical significance of the *calibration* results. To obtain tests, one must resort to an *estimation* procedure which requires making reasonable assumptions on the distribution of the stochastic term e_s. A major difficulty in this respect is the fact that, for an arbitrarily given value of λ, one is almost certain that σ^2, the presumed variance of this error term, is not constant because large errors will occur on links with much traffic and small errors on other links. An assumption of constant variance would then lead to unreliable estimates of standard errors and dubious tests of significance. In our network, dominated by the Montreal metropolitan area, the presence of heteroskedasticity would typically imply, for the linear case of $\lambda = 1$, underestimating the standard errors of the β coefficients. A way out of this predicament would consist of formulating the model in such a way that the resulting error variance would be constant. This is currently feasible for the linear model ($\lambda = 1$) but not for the more general specification in (5). Yet it is known that the parameter λ may be very sensitive to the presence of heteroskedasticity (Zarembka, 1974), a fact which has been confirmed (Gaudry and Wills, 1977) in an intercity travel demand model of type (3). The

(5). Yet it is known that the parameter λ may be very sensitive to the presence of heteroscedasticity (Zarembka, 1974), a fact which has been confirmed (Gaudry and Wills, 1978) in an intercity travel demand model of type (3). The parameter λ "absorbs" heteroscedasticity and confounds functional form and random term specification issues. Until a joint formulation of heteroscedasticity and functional form has been developed and tested, the choice then seems to be between unreliable estimates of functional form in the presence of heteroscedasticity and reasonable corrections for heteroscedasticity under linearity assumptions.

C. A linear specification with homoscedastic errors

The chosen model is

$$F_s = \sum_i \sum_j p_{ij}^s \left[\beta_o + \sum_k \beta_k X_{ij_k} \right] + e_s \qquad (6\text{-A})$$

and a variant in which the constant term is constrained equal on all links:

$$F_s = \beta_o + \sum_i \sum_j p_{ij}^s \left[\sum_k \beta_k X_{ij_k} \right] + e_s \quad . \qquad (6\text{-B})$$

The latter does not allow for an estimation of the complete O-D matrix because it provides no systematic way of associating the constant with specific O-D pairs: this makes no difference for forecasting purposes. In both variants, the covariance matrix is assumed to be of the form

$$E(e_s e_s') = \sigma^2 \Omega = \sigma^2 \begin{bmatrix} [f(X_{k*})_1]^2 & & \bigcirc \\ & \ddots & \\ \bigcirc & & [f(X_{k*})_n]^2 \end{bmatrix}, \qquad (7)$$

where e_s denotes the n x 1 vector of link disturbances, k* denotes population, the variable most likely to influence error variance, and $f(X_{k*}) = \sum_i \sum_j p^s_{ij} X_{ij_{k*}}$.

We applied the parametric test proposed by Goldfeld and Quandt (1965) using, in Equation 6, the average daily flow of traffic in 1972 as dependent variable and, in addition to travel time, the following explanatory variable specification:

$$X_{ij_k} = X_{i_k} + X_{j_k} \tag{8}$$

for population, number of manufacturing jobs, car ownership, average household income and the proportion of French-speaking population. The resulting value of the test statistic was 7.13 which implied that one could reject with an extremely high degree of confidence the null hypothesis of homoscedasticity (the critical value of the F-distribution at the 5% level was 2.3).

We then used the Aitken generalized least-squares estimator

$$b = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}y$$

to explain both average summer and non-summer traffic flows with various lists of explanatory variables. The results for the latter flows were not significantly different from the results for the former (Lamarre, 1977) which will be reported here. The income and linguistic index variables were also removed because of their high correlation (> .99) with the constant term in (6-A).

3. RESULTS

A. General comments

The principal results for both model variants are shown in Table 1. The following comments can be made:

- the constant and travel time are generally the only statistically significant variables;
- the results for both variants are mediocre;
- travel time demand elasticities, shown in Table 2, are more convincing for variant 6-B than for variant 6-A. The latter are all above 2 and imply that a 10% change in travel speeds would increase or decrease traffic volumes by about 23% which is absurd;
- it is clear from an examination of residuals on links (not shown) that computed flows are systematically overestimated for peripheral links and adequately estimated close to Montreal. This suggests a non-linear impact of distance on the level of trip making.

B. Summary comparison with gravity format results

In order to obtain an idea of the possible presence of non-linearities, a generation-distribution gravity form of (5) was *calibrated* using, as in variant 6-A (iii), a constant, population and distance as explanatory variables. The calibrated coefficients were respectively: 2.48, 0.67 and - 1.62. The computations were made with a non-linear optimization routine (Wills, 1974) which minimizes the sum of squared errors in (5) without correcting for heteroskedasticity.

TABLE 1

RESULTS FOR LINEAR SPECIFICATION*

Variables	Variant 6-A			Variant 6-B		
	i	ii	iii	i	ii	iii
Constant	1725 (2.08)	1804 (8.93)	1802 (9.22)	1753 (0.27)	2858 (0.97)	2797 (1.91)
Population	-0.0002 (-0.04)	-0.00002 (-0.00)	-0.002 (-0.36)	0.001 (0.27)	0.004 (0.97)	0.001 (1.91)
% manufact. jobs	0.37 (0.10)			4.00 (2.80)		
Car ownership	-0.0002 (-0.01)	-0.0007 (-0.05)		-0.003 (-0.22)	-0.01 (-0.71)	
Travel time	-4.42 (-2.63)	-4.56 (-5.16)	-4.55 (-5.24)	-1.12 (-1.73)	-0.74 (-1.06)	-0.77 (-1.11)
R-SQUARED	.72	.72	.72	.77	.71	.71

*Numbers in parentheses are t-statistics.

TABLE 2

TRAVEL TIME DEMAND ELASTICITIES*

LINEAR FORM						GRAVITY FORM
Variant 6-A			Variant 6-B			Variant 6-A
i	ii	iii	i	ii	iii	iii
-2.24	-2.31	-2.31	-0.57	-0.38	-0.39	-1.62

*Computed at the mean.

It was noted that

- the resulting R^2 was .88, which is higher than for the linear model and could imply the existence of non-linearities if it were robust to heteroskedasticity;
- as shown in Table 2, the very high travel time elasticity of demand obtained is close to the value for the corresponding linear model but slightly more reasonable.

4. CONCLUSION

These results are sufficiently interesting to suggest that the estimation of O-D matrices from link counts is a promising and inexpensive way of analyzing traffic data if progress can be made on the problem of the joint determination of multi-element cost matrices and on the adequate simultaneous estimation of heteroskedasticity and functional form parameters in order not to confound functional form issues with stochastic mispecification: until econometric methods have progressed in this direction, a rigorous comparison of linear and multiplicative specifications will not be possible.

ACKNOWLEDGEMENTS

The research on which this article is based was partly supported by the Research and Development Centre of Transport Canada and by the F.C.A.C. program of the *Ministère de l'éducation du Québec*. Most of the work was done in 1975 and 1976. The first author is associated with the *Département de sciences économiques* and the *Centre de recherche sur les transports* of *L'Université de Montréal*, the second author is employed by LAVALIN, Inc. They both thank the *Ministère des transports du Québec* for supplying data and Michael J. Wills of the Strategic Planning Group of Transport Canada, Ottawa, for using his program. They are also grateful to Wills for supplying and permitting references to unpublished material.

REFERENCES

Arbour, D., Bégin, P., Gohier, J.-P., Latour, G. et Thivierge, J. (1969). Le modèle de transport: textes polycopiés présentés au Comité AMUR-ETEN. Service d'urbanisme. Ville de Montréal

Arbour, D., Bégin, P., Gohier, J.-P., Latour, G. et Thivierge, J. (1971a). Le modèle de transport: schéma. Service d'urbanisme. Ville de Montréal.

Arbour, D., Bégin, P., Gohier, J.-P., Latour, G. et Thivierge, J. (1971b). Le modèle de transport: simulations métro 1981. Service d'urbanisme. Ville de Montréal.

Bendtsen, P.H. (1974). Simplification of traffic models. Traffic Engineering and Control 15. 817-819.

Box, G.E.P. and Cox, D.R. (1964). An analysis of transformations. Journal of the Royal Statistical Society Series B, 211-243.

Branston, D., Florian, M., James, L., Nguyen, S., Trahan, M. (1975). Développement d'un outil d'étude pour déterminer l'impact de nouveaux aménagements sur la circulation routière de l'île de Hull. Publication #46. Centre de recherche sur les transports. Université de Montréal.

Debaille, S. (1977). Un modèle de reconstitution de matrices origine-destination en milieu urbain. Rapport no. 27. Institut de recherche des transports. Arcueil.

Gaudry, M.J.I. (1977). Six notions of equilibrium and their implications for travel modelling examined in an aggregate direct demand framework. Publication #51. Centre de recherche sur les transports. Université de Montréal. Forthcoming in Hensher, D. and Stopher, P., editors of Behavioural Travel Modelling, Croom Helm, London, ch. 6.

Gaudry, M.J.I. and Wills, M.J. (1977). Estimating the functional form of travel demand models. Publication #63. Centre de recherche sur les transports. Université de Montréal. Forthcoming in Transportation Research.

Goldfeld, S.M. and Quandt, R.E. (1965). Some tests for homoskedasticity. Journal of the American Statistical Association 60. 539-47.

Högberg, P. (1976). Estimation of parameters in models for traffic prediction: a non-linear regression approach. Transportation Research 10. 263-265.

Holm, J., Jensen, T., Nielsen, S.K., Christensen, A., Johnsen, B. and Ronby, G. (1976). Calibrating traffic models on traffic census results only. Traffic Engineering and Control 17. 137-140.

REFERENCES . . .

Jentsen, T. and Nielsen, S.K. (1973). Calibrating a gravity model and estimating its parameters using traffic volume counts. Paper presented to the 5th Conference of Universities Transport Study Group, University College London.

Lamarre, L. (1977). Une méthode linéaire simple d'estimation de la matrice des origines et des destinations à partir de comptages sur les liens d'un réseau: une application au réseau routier du Québec. Publication #76. Centre de recherche sur les transports. Université de Montréal.

Low, D. (1972). A new approach to transportation system modeling. Traffic Quarterly 26. 391-404.

Nguyen, S. (1977a). Estimating an O-D matrix from network data: a network equilibrium approach. Publication #60. Centre de recherche sur les transports. Université de Montréal.

Nguyen, S. (1977b). On the estimation of an O-D trip matrix by the equilibrium methods using pseudo delay functions. Publication #81. Centre de recherche sur les transports. Université de Montréal.

Nguyen, S. (1978). On estimating origin-destination matrices from traffic data. Paper presented at the Canadian Transportation Research Forum Meeting, Winnipeg. June 12-14.

O.E.C.D. Road Research Group (1974). Urban Traffic Models: Possibilities for simplification. O.E.C.D. Paris.

Overgaard (1972). Development of a simplified traffic model for the city of Silkeborg. O.E.C.D. seminar in Copenhagen.

Robillard, P. and Trahan, M. (1973). Estimating the O-D matrix and network characteristics from observed link volumes. Publication #128. Département d'informatique. Université de Montréal. Proceedings of the International Conference on Transportation Research. First Conference. 736-740.

Robillard, P. (1974). Estimating the O-D matrix from observed link volumes. Publication #1. Centre de recherche sur les transports. Université de Montréal. Transportation Research 9. 123-128. 1975.

Van Zuylen (1978). A method to estimate a trip matrix from traffic volume counts. PTRC Summer Annual Meeting.

Wills, M.J. (1971). Development of the highway network, traffic flow and the growth of settlements in interior British Columbia. M.A. Thesis, Department of Geography, The University of British Columbia, Vancouver.

REFERENCES . . .

Wills, M.J. (1974). Spatial linkages submodel: inter-urban passenger movement project: models and computer programs. Ministry of State for Urban Affairs, Ottawa.

Willumsen, L. (1978a). Estimation of an O-D matrix from traffic counts: a review. Working Paper No. 99. Institute for Transport Studies. Leeds University.

Willumsen, L. (1978b). O-D matrices from network data: a comparison of alternative methods for their estimation. PTRC Summer Annual Meeting.

Zarembka, P. (1974). Transformation of variables in econometrics. Ch. 3 of Frontiers in Econometrics, ed. by P. Zarembka.